



# Attention-based deep convolutional neural network for spectral efficiency optimization in MIMO systems

Danfeng Sun<sup>1,2</sup> · Abdullah Yaqot<sup>2</sup> · Jiachen Qiu<sup>1</sup> · Lutz Rauchhaupt<sup>2</sup> · Ulrich Jumar<sup>2</sup> · Huifeng Wu<sup>1</sup>

Received: 6 April 2020 / Accepted: 16 June 2020 / Published online: 9 July 2020  
© Springer-Verlag London Ltd., part of Springer Nature 2020

## Abstract

Spectral efficiency (SE) optimization in massive multiple input multiple output (MIMO) antenna cognitive systems is a challenge originated from the coexistence restrictions. Traditional power allocation can optimize the SE; however, involving deep learning can meet real-time and fairness processing requirements. In unfair allocation problem, all power is possibly assigned to one or few antennas of a particular user. In this paper, we build a mathematical optimization model considering the fairness problem such that SE is optimized for all users. To implement the model, we propose an attention-based convolutional neural network (Att-CNN), where  $h_0$  and  $h_k$  (i.e., cross-interference and direct channels) attention mechanisms are used to improve the SE. The convolutional neural network is applied to decrease the floating point operations (FLOPs) and number of network parameters. We conducted experiments from these aspects: Fair antenna power allocation, power allocation performance and computational performance. Heat maps with different interference thresholds show the fair allocation for all users. Analyses of SE validate the superiority of the Att-CNN compared with the equal power allocation and fully connected neural network (FNN) schemes. The analyses of the FLOPs and number of parameters show the superiority of the Att-CNN over the FNN.

**Keywords** Attention mechanism · Cognitive radio · Convolutional neural network · Massive MIMO system · Spectral efficiency optimization

## 1 Introduction

Big data communications in high connection density networks are ever-growing because of the increasing demands in social networks [1, 2], augmented reality [3], etc., under the background of the Internet of Things. Spectral efficiency optimization can enhance the throughput to ease this problem. This motivates novel resource allocation designs in multiple input multiple output (MIMO) cognitive radio (CR) to address such high data rate applications. Cognitive radio has been known as a technology for reutilizing the spectrum licensed for a primary radio (PR) [4]. In the underlay CR mode, both PR and CR can concurrently

transmit without inducing harmful interference by CR transmitters at PR receivers in that the CR interference power at PR is restricted below a predefined threshold [5, 6].

At the heart of today's wireless networks, e.g., IEEE-based [7] and 3GPP-based [8], the combination of orthogonal frequency division multiplexing with the multiple input multiple output technologies is a feature for addressing physical (PHY) and medium access control (MAC) layers challenges. Consequently, the communication system possesses large number of radio resources and degrees of freedom for which the management processing demands low complexity algorithms to fulfill requirements of emerged applications and use cases. This is a challenging aspect in the conventional algorithms especially for the CR networks. For convenience, minimum mean square error (MMSE) algorithm [9] demands complex operations such as matrix inversion in every iteration. Interior point algorithm based on barrier method [10] employs centric iterations in which complex Newton step is executed.

✉ Huifeng Wu  
whf@hdu.edu.cn

<sup>1</sup> Institute of Industrial Internet, Hangzhou Dianzi University, Hangzhou 310018, China

<sup>2</sup> Institut f. Automation und Kommunikation, Magdeburg 39106, Germany

Iterative waterfilling algorithm [11] involves singular value decomposition at each iteration. Lagrange algorithms [12, 13] solve closed-form expressions in terms of iterative dual variables associated with the optimization constraints. Generally speaking, the aforementioned iterative optimization algorithms are by nature computationally expensive. That makes their real-time implementation challenging especially for large-scale wireless communication system whose algorithm execution is periodically required in time frames of milliseconds (due to the very dynamic system parameters such as channel conditions, number of users, etc.).

On the other hand, machine learning has been a research trend in several processings fields of wireless communications such as resource and interference management [14, 15], channel state information feedback and estimation [16, 17], antenna selection and beamforming design [18, 19] and others. The reason behind is the potential computational efficiency of deep learning solvers in optimization problems.

Hence, this paper proposes an attention-based deep learning algorithm to address a novel power allocation design in CR MIMO systems. Our contributions are summarized as follows.

1. We developed a fair per-antenna power allocation scheme in CR MIMO systems. The proposed scheme addresses the issue of opportunistic unfair power allocation in that all users can have the basic quality of service (QoS)<sup>1</sup> by setting lower bounds during the configuration.
2. We propose an attention-based convolutional neural network (Att-CNN) method to implement the fair power allocation over antennas and users. We design two types of attention mechanisms: Cross-channel  $h_0$  based and direct channel  $h_k$  based, which support the whole neural network to focus on the relationships between  $h_0$  and  $h_k$ , and inside  $h_k$ , respectively. In addition, the CNN effectively reduces the floating point operations (FLOPs) and number of network parameters.
3. Beside the implementation of Att-CNN model, we also implement the proposed per-antenna power allocation paradigm using fully connected neural network and equal power allocation methods. Main findings show that the Att-CNN outperforms those baselines.

Throughout the paper,  $(\cdot)^T$  and  $(\cdot)^H$  are the transpose and conjugate transpose, respectively.  $\|\cdot\|$  denotes the 2-norm.

<sup>1</sup> Basic Qos in this sense means the ability to get the communication service with a data rate proportional to channel conditions. Satisfactory QoS via securing a minimum data rate goes along with the CR spectral efficiency enhancement goal only when good channel conditions are witnessed.

$[\mathbf{X}]_i$  denotes the  $i$ th column of the matrix  $\mathbf{X}$ . For real value  $x$ ,  $[x]^+ = \max(x, 0)$ .  $\mathbb{R}^{m \times n}$  denotes the space of real  $m \times n$  matrices, respectively.

The sequel of this paper is organized as follows: Sect. 2 exhibits the system model and the problem statement. Section 3 derives the proposed fair per-antenna power allocation by means of the Att-CNN. Section 4 presents the experimental results. Section 5 introduces the related work. Finally, Sect. 6 concludes the paper.

## 2 Problem statement

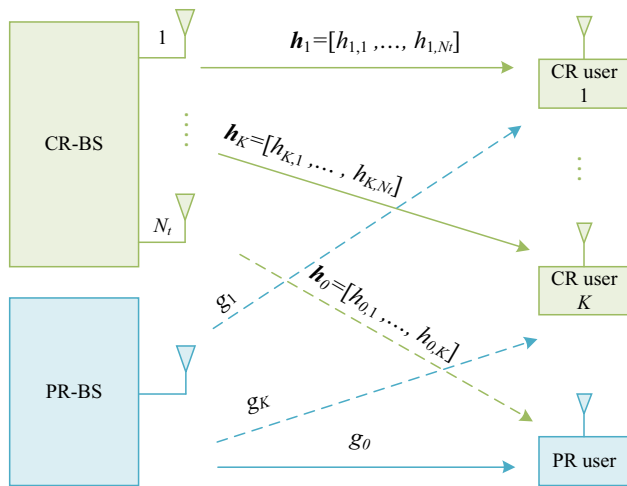
We consider a CR network including a CR base station (CR-BS) and coexisting with a PR base station (PR-BS). The CR-BS is equipped with  $N_t$  antennas communicating with  $K$  single-antenna CR users (CUs) where  $N_t, K \in \mathbb{R}^+$ . The PR-BS is a single antenna system serving a primary user. The antenna system of the CR-BS is a massive MIMO. Mutual interference is assumed according to the underlay CR setting in which CR induces interference at the primary user but below a predefined threshold, while the PR-BS introduces interference at the CR users without restrictions. The channels between the CR-BS antenna  $i$  and CR user  $k$ , the CR-BS antenna  $i$  and the PR user, the PR-BS and the CR user  $k$  and the PR-BS and PR user are denoted as  $h_{k,i}$ ,  $h_{0,i}$ ,  $g_k$  and  $g_0$ , respectively, where  $1 \leq k \leq K$  and  $1 \leq i \leq N_t$ . The total power of PR-BS and CR-BS are denoted as  $P_{PR}$  and  $P_t$ , respectively. The noise power and interference threshold caused by the CR-BS at the PR user are defined as  $\sigma$  and  $I_{th}$ . System model is illustrated in Fig. 1, and all parameters are listed and explained in Table 1.

In this paper, we intend to optimize the spectral efficiency of CR user  $k$  in the proposed underlay CR network, which can be written as follows:

$$SE_{CR}^k = \log_2 \left( 1 + \frac{\| \sum_{i=1}^{N_t} h_{k,i} P_{k,i}^{1/2} \|^2}{\sigma^2 + \sum_{l \neq k} \| \sum_{i=1}^{N_t} h_{k,i} P_{l,i}^{1/2} \|^2 + \xi} \right), \tag{1}$$

where  $\xi = \| g_k \|^2 P_{PR}$  is the interference from the PR-BS.

Intuitively, the optimization problem of antennas power allocation can be stated as follows:



**Fig. 1** System Model: MIMO CR network coexists with a PR network

$$\begin{aligned}
 \mathcal{SE}_{CR} &= \max_{\{P_{k,i} \geq 0\}} \sum_{k=1}^K \mathcal{SE}_{CR}^k, \\
 \text{s.t. :} \\
 C1 : & \sum_{k=1}^K \sum_{i=1}^{N_t} P_{k,i} \leq P_t, \\
 C2 : & I_{CR} = \sum_{k=1}^K \left\| \sum_{i=1}^{N_t} h_{0,i} P_{k,i}^{1/2} \right\|^2 \leq I_{th}.
 \end{aligned} \tag{2}$$

Here, the optimizing target is to maximize the SE of all cognitive users  $\mathcal{SE}_{CR}$  in the CR network, under the constraints of total power  $P_t$  and interference threshold  $I_{th}$ . However, when we addressed this optimization problem with tuned deep neural networks, we met a new problem. Figure 2 shows a heat map example for unfair power allocation. The horizontal and vertical axes are users and antennas indices, respectively. It is clear that a single CU or

even several antennas of that user are allocated almost all the power. This is unfair to other users, because all have the same priority. The fair scheme involves more spatial degrees of freedom in the MIMO system via serving more users, i.e., multiuser MIMO. However, the unfair schemes engage single user underutilizing the spatial degrees of freedom, i.e., single-user MIMO, given that the antenna condition  $N_t \geq K$  holds [9, 12]. Thus, the benefit is manifold: better QoS and higher spectral efficiency performance. It is worth noting that the global maximum solution can only be found by an exhaustive search, which is based on trial-and-error method whose computational complexity is inevitably intractable for real-time communications.

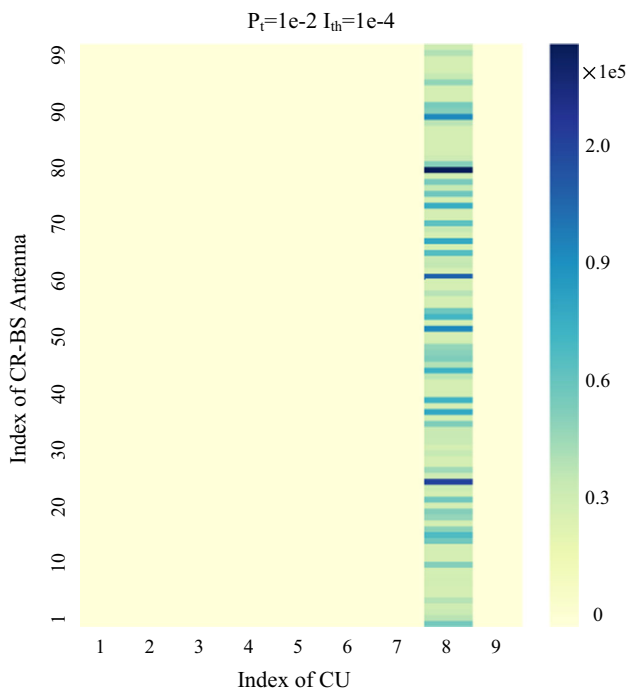
Therefore, we anchor the optimization problem to fair antennas power allocation and add a new constraint (i.e., user minimum power  $\lambda_k$ ) as follows:

$$\begin{aligned}
 \mathcal{SE}_{CR} &= \max_{\{P_{k,i} \geq 0\}} \sum_{k=1}^K \mathcal{SE}_{CR}^k, \\
 \text{s.t. :} \\
 C1 : & \sum_{i=1}^{N_t} P_{k,i} \leq \lambda_k, \\
 C2 : & \sum_{k=1}^K \sum_{i=1}^{N_t} P_{k,i} \leq P_t, \\
 C3 : & I_{CR} = \sum_{k=1}^K \left\| \sum_{i=1}^{N_t} h_{0,i} P_{k,i}^{1/2} \right\|^2 \leq I_{th},
 \end{aligned} \tag{3}$$

where  $\lambda_k$  is configurable. Note that if  $\lambda_k = 0, \forall k \in [1, K]$ , then it becomes the previous unfair antennas power allocation problem.

**Table 1** Definitions of system model parameters

Parameter	Definition
$\mathcal{SE}_{CR}^k$	Spectral efficiency of CR user $k$
$\mathcal{SE}_{PR}$	Spectral efficiency of PR user
$h_{k,i}$	Channel gain between CR-BS antenna $i$ and CR user $k$
$h_{0,i}$	Channel gain between CR-BS antenna $i$ and primary user $k$
$g_k$	Channel gain between the PR-BS and cognitive user $k$
$g_0$	Channel gain between the PR-BS and primary user
$P_{k,i}$	Allocated power to antenna $i$ for user $k$
$\sigma^2$	Noise power
$P_{PR}$	Total power of PR-BS
$P_t$	Total power of CR-BS
$I_{CR}$	Interference from the CR network to the PR user
$I_{th}$	Interference threshold at the primary user
$N_t$	Number of CR-BS antennas
$K$	Number of CR users



**Fig. 2** Heat map of antennas power allocation. It shows an example for unfair power allocation scheme that assigns all the power to a single CR user (8th index)

### 3 Attention-based deep neural network

#### 3.1 Network structure

We use an attention-based convolutional neural network, i.e., Att-CNN, to address the above-mentioned problem. Because the channel gains obtained from our channel model are extremely small, we use the following equation to normalize the dataset.

$$\hat{h}_{i,j} \triangleq \frac{\log_{10}(h_{i,j}) - \mathbb{E}[\log_{10}(h_{i,j})]}{\sqrt{\mathbb{E}[(\log_{10}(h_{i,j}) - \mathbb{E}[\log_{10}(h_{i,j})])^2]}} \tag{4}$$

To process the complex numbers of channel gains, a data initialization network is used, see Fig. 3. We define the block channel matrix as  $\mathbf{H} \triangleq [\mathbf{h}_1^T, \dots, \mathbf{h}_k^T]^T$ . The normalized channel gains of  $\mathbf{H}$  are input of  $N_i$  full connected layers (FCLs) with  $L_i$  cells to memorize the relationship between real and imaginary parts. Then a cross-channel  $h_0$  attention network accepts the data and outputs a new matrix called  $\hat{\mathbf{H}}$ , which is mixed real and imaginary parts of channel gains and adjusted by cross-channel  $h_0$ . The cross-channel  $h_0$  attention network is introduced in Sect. 3.2.1.

Then the integrated matrix is input into three networks, see Fig. 3. At the former of every network is a direct channel  $h_k$  attention network (HKAN), which is used to

reevaluate every row in the integrated matrix, see Sect. 3.2.2.

In Network 1, the output of the HKAN is input into  $N_c$  CNN layers. The CNN can decrease the number of neural network parameters because of parameter sharing. Pooling layers are not considered since all data should be used.  $N_c$  is limited by the receptive field that the receptive field of the top most layer should be no larger than the input image region [20], which can be calculated as follows.

$$r_n = r_{n-1} * l_n - (l_n - 1) * \left( r_{n-1} - \prod_{i=1}^{n-1} s_i \right), s.t., n \geq 2, \tag{5}$$

where  $l_n$   $s_n$   $r_n$  are kernel size, stride and receptive field of  $n$ th layer, respectively;  $r_0 = 1$ ; and  $r_1 = l_1$ . Figure 4 is an example that the input image size is  $10 * 10$  pixels; and the kernel size and stride of the first and second layer both are  $4 * 4$  pixels and  $1 * 1$  pixel, then the receptive field of second layer is  $7 * 7$  pixels. At last, the receptive field of the third layer reaches  $10 * 10$  pixels, which covers the whole image. The output of the CNN layers is sent into  $N_{f1}$  FCLs with  $L_{f1}$  cells, and with a softmax layer, we obtain the final output of Network 1 which is seen as  $\frac{P_{k,i}}{P_k}$ , because it has the same mathematical characteristic with softmax as follows.

$$\sum \frac{P_{k,i}}{P_k} = 1. \tag{6}$$

In Network 2, the output of the HKAN is directly input into  $N_{f2}$  FCLs with  $L_{f2}$  cells. Then a FCL with sigmoid activation functions is used to produce the final output of Network 2, and this output is denoted as  $\frac{P'_k}{P'_k}$ , where  $P'_k$  is the upper bound of  $P_k$ , and  $\frac{P'_k}{P'_k}$  has the same range with sigmoid as follows.

$$0 < \frac{P'_k}{P'_k} < 1. \tag{7}$$

Similar with Network 2, the output of the HKAN in Network 3 is input into  $N_{f3}$  FCLs with  $L_{f3}$ . Then the output is fed into a softmax and a sigmoid layer, respectively. The softmax and sigmoid layers output  $\frac{P'_k}{\sum P'_k}$  and  $\frac{\tilde{P}_k}{P_T}$ , respectively, because they have the same ranges with softmax and sigmoid as follows.

$$\sum \frac{\tilde{P}_k}{\sum \tilde{P}_k} = 1, \tag{8}$$

$$0 < \frac{\sum \tilde{P}_k}{P_T - \sum \lambda_k} < 1, \tag{9}$$

where  $\tilde{P}_k$  is the up bound power which can allocate to user  $k$  beside  $\lambda_k$ . Namely,  $P'_k = \lambda_k + \tilde{P}_k$ .

Then we can calculate  $P_{k,i}$  as follows:

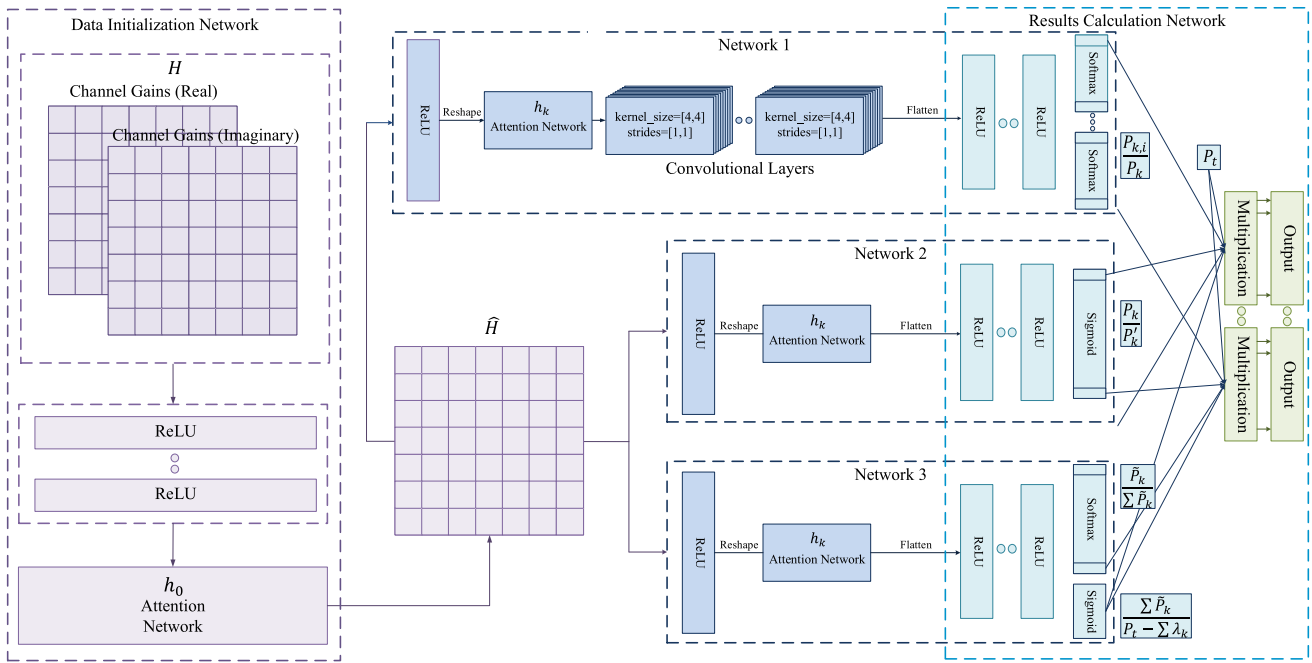


Fig. 3 Structure of the proposed attention-based convolutional neural network

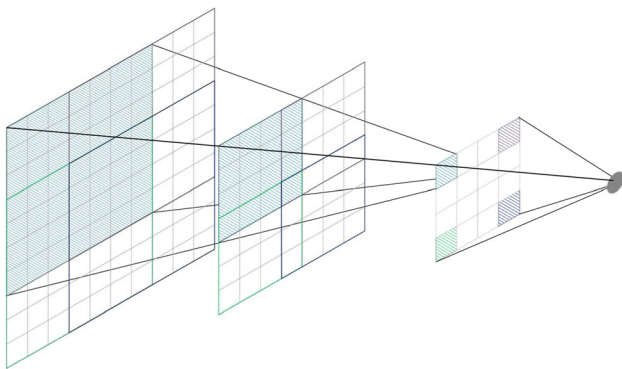


Fig. 4 Receptive field

$$P_{k,i} = \frac{P_{k,i} P'_k}{P_k P'_k} \left[ \lambda_k + \frac{\tilde{P}_k}{\sum \tilde{P}_k} \frac{\sum \tilde{P}_k}{P_T - \sum \lambda_k} (P_T - \sum \lambda_k) \right] \quad (10)$$

Thus, we obtain  $P_{k,i}$  which complies with the constraints of C1 and C2. Then, regarding our fair antennas power allocation problem, loss function is designed as follows:

$$Loss \triangleq - \sum_{k=1}^K \log_2 \left( 1 + \frac{\| \sum_{i=1}^{N_i} h_{k,i} \hat{P}_{k,i}^{1/2} \|^2}{\sigma^2 + \sum_{l \neq k} \sum_{i=1}^{N_l} \| h_{k,i} \hat{P}_{l,i}^{1/2} \|^2 + \xi} \right), \quad (11)$$

where  $\hat{P}_{k,i}$  is calculated by Algorithm 1. Note that  $\hat{P}_{k,i} = P_{k,i} / (1 + [ICR/I_{th} - 1]^+)$  which makes a penalty to  $P_{k,i}$  when the constraint C3 is violated. However, apparently,  $\hat{P}_{k,i}$  must not be bigger than  $P_{k,i}$  which may not meet the

constraint C1 in some cases. Thus, in order to find the equilibrium between constraint C1, C2 and C3, we propose the iterative Algorithm 1.

**Algorithm 1:** Iteration for Constraints

**Input:**  $P_{k,i}, N_t$

**Output:**  $\hat{P}_{k,i}$

$\hat{P}_{k,i} = P_{k,i};$

**for** 0 to  $N_d$  **do**

$$\hat{P}_{k,i} = \frac{\hat{P}_{k,i}}{(1 + [ICR/I_{th} - 1]^+)};$$

$$\hat{P}_{k,i} = \frac{\hat{P}_{k,i}}{(1 - [1 - \sum_i \hat{P}_{k,i} / \lambda_k]^+)};$$

$$\hat{P}_{k,i} = \frac{\hat{P}_{k,i}}{(1 + [P_t / \sum_k \sum_i \hat{P}_{k,i} - 1]^+)};$$

**end**

**3.2 Applications of attention mechanism**

Attention mechanism is a technique imitating human beings to address problem which focuses on important information from big data. In our problem, on account of the diversity of channel gains  $H$ , the networks may get wrong rules, which should be nonexistent, learning from the limited training sets. The proposed attention mechanism is specific and makes the networks more sensitive to channel gains  $H$  from its unique structure compared to fully connected networks. We define two kind of attention mechanism:  $h_0$  and  $h_k$ , which are represented as two neural

networks. The cross-channel  $h_0$  attention network takes into account the relationship between the  $h_0$  and  $h_k$ , and the direct channel  $h_k$  attention network represents the internal relationship of  $h_k$ . The details are reported as follows.

### 3.2.1 The cross-channel $h_0$ attention network

As shown in Eq. (3), the channel gain between CR antenna  $i$  and the PR user  $h_0$  can influence the value of  $I_{CR}$ , which is included in constraint C3. Hence, the cross-channel  $h_0$  attention mechanism is involved in the initialization network. In other words, we reevaluate the weights of CR channel gains in terms of  $h_0$ . Figure 5a shows the cross-channel  $h_0$  attention network. Here,  $h_0$  is used to produce a matrix  $q \in \mathbb{R}^{\bar{K} \times N_t}$ ,  $\bar{K} = K/2$  as follows:

$$q = \sigma(W^q h_0), \tag{12}$$

where  $W^q$  is a  $\bar{K} \times 1$  weight vector from training, and  $\sigma$  is an activation function. Note that  $\bar{K} = K/2$  is proposed for reducing computation. In addition, matrix  $f \in \mathbb{R}^{\bar{K} \times N_t}$  is taken from CR channel gain matrix  $H$  as follows:

$$f = \sigma(W^f H), \tag{13}$$

where  $W^f$  is the  $\bar{K} \times K$  parameter matrix. Then, matrix  $\alpha \in \mathbb{R}^{N_t \times N_t}$  is defined as the extent to which the network attends to  $q$  when adjusting  $f$ .  $\alpha_{ij}$  can be obtained as follows:

$$\alpha_{ij} \triangleq \frac{\exp(\text{score}_{ij})}{\sum_{i=1}^{N_t} \exp(\text{score}_{ij})}, \tag{14}$$

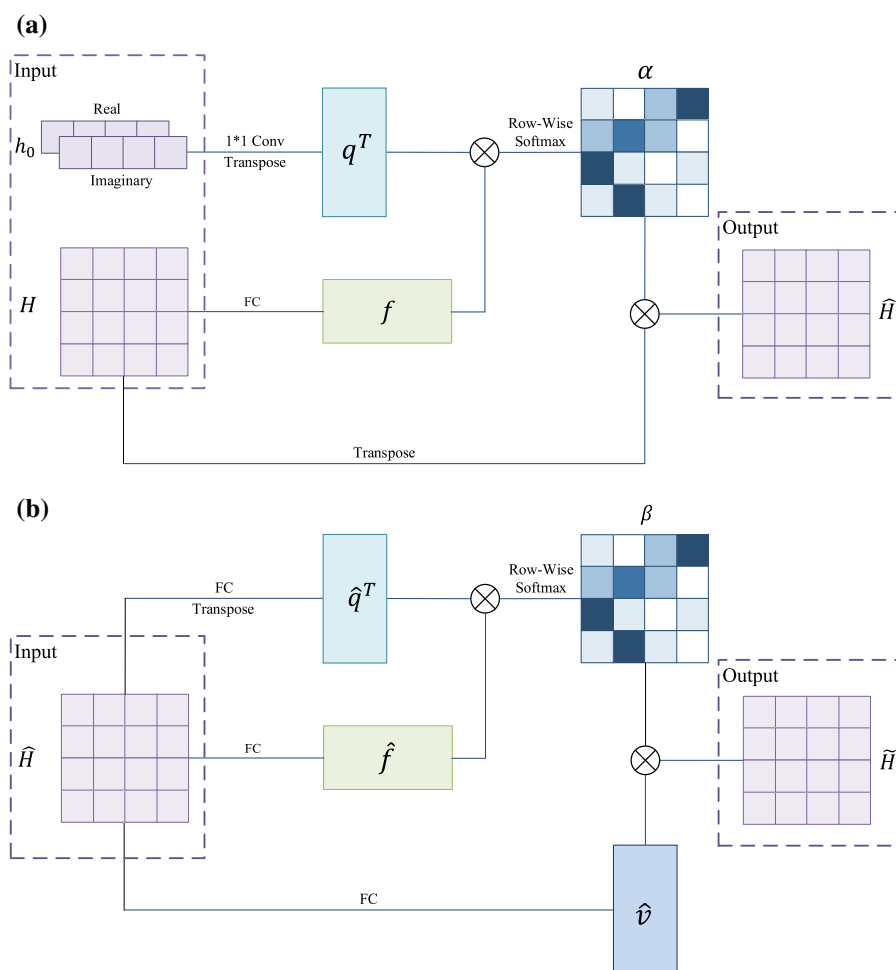
where  $\text{score} = q^T f \in \mathbb{R}^{N_t \times N_t}$ . Then, we can obtain a new channel gain matrix with the cross-channel  $h_0$  attention mechanism written as  $\hat{H} \triangleq [\hat{h}_1^T, \hat{h}_2^T, \dots, \hat{h}_j^T, \dots, \hat{h}_K^T]^T \in \mathbb{R}^{K \times N_t}$ , where

$$\hat{h}_j = h_j \alpha. \tag{15}$$

### 3.2.2 The direct channel $h_k$ attention network

Equation (1) indicates that its value is relevant to  $h_{k,i} P_{k,i}^{1/2}$  and  $h_{l,i} P_{k,i}^{1/2}$ . This means the channel gain relationship

**Fig. 5** **a** The cross-channel  $h_0$  attention network. **b** The direct channel  $h_k$  attention network



among users can influence the result of  $\mathcal{SE}$ . Therefore, we design the direct channel  $h_k$  attention network. Note that because of the different functions of Network 1, 2 and 3, they do not share the same direct channel  $h_k$  attention network. In addition, for Network 1, individually using CNN ignores the relationships between non-adjacent rows in the  $\mathbf{H}$ , the direct channel  $h_k$  attention network can support CNN to provide a better performance. The network structure of  $h_k$  attention network is similar  $h_0$  attention network. Noted that, the input of the network is  $\hat{\mathbf{H}}$  (which gets off  $h_0$  attention network) and matrix  $\hat{\mathbf{q}}$  is produced from  $\hat{\mathbf{H}}$  itself instead of  $\mathbf{h}_0$ . In addition, we introduce a new matrix  $\hat{\mathbf{v}}$ . The matrices  $\hat{\mathbf{q}} \in \mathbb{R}^{K \times N_t}$ ,  $\hat{\mathbf{f}} \in \mathbb{R}^{K \times N_t}$ ,  $\hat{\mathbf{v}} \in \mathbb{R}^{K \times N_t}$  can be described as follow:

$$\begin{cases} \hat{\mathbf{q}} = \sigma(\hat{\mathbf{W}}^q \hat{\mathbf{H}}), \\ \hat{\mathbf{f}} = \sigma(\hat{\mathbf{W}}^f \hat{\mathbf{H}}), \\ \hat{\mathbf{v}} = \sigma(\hat{\mathbf{W}}^v \hat{\mathbf{H}}), \end{cases} \quad (16)$$

where  $\hat{\mathbf{W}}^q$ ,  $\hat{\mathbf{W}}^f$ ,  $\hat{\mathbf{W}}^v$  are  $K \times K$  weight matrices. Similar to Equation (14),  $\beta_{i,j}$  can be obtained by the inner product of extracted vectors (16) as follows:

$$\beta_{i,j} \triangleq \frac{\exp([\hat{\mathbf{q}}]_i^T [\hat{\mathbf{f}}]_j)}{\sum_{i=1}^{N_t} \exp([\hat{\mathbf{q}}]_i^T [\hat{\mathbf{f}}]_j)}. \quad (17)$$

where  $[\hat{\mathbf{q}}]_i$  and  $[\hat{\mathbf{f}}]_j$  are the  $i$ th and  $j$ th columns of  $\hat{\mathbf{q}}$  and  $\hat{\mathbf{f}}$ , respectively. We define  $\mathbf{H} \triangleq [\mathbf{h}_1^T, \mathbf{h}_2^T, \dots, \mathbf{h}_j^T, \dots, \mathbf{h}_K^T]^T \in \mathbb{R}^{K \times N_t}$  as the output of the direct channel  $h_k$  attention network, where

$$\tilde{\mathbf{h}}_j = \mathbf{v}_j^T \boldsymbol{\beta}. \quad (18)$$

## 4 Experiment

We built a channel model in terms of [21] to produce channel gains, which takes path loss and multi-path fading effects into account. The following configuration is used. Path loss exponent is set 2.5, distance between CUs/PU and CR/PR base stations is a uniformly distributed random variable in the range [10, 200]. Using this channel model, we produced a data set which has 1000  $10 \times 100$  matrices for the following Sects. 4.1 and 4.2. We assumed  $N_t = 99$  antennas and  $K = 9$  CR users, and in every  $10 \times 100$  matrix, the  $h_{k,i}$  elements occupy the first 9 rows and 99 columns; the  $h_{0,i}$  elements occupy the last row index (column index 1 to 99); the  $g_k$  elements occupy the last column index (row indices 1 to 99) and the element of  $g_0$

occupies the last column index and last row index. 90% of the data set was used for training, 10% for testing. Such training/testing is well used in data mining and machine learning domains [22–24]. The noise was generated as circular symmetric complex Gaussian random variable with zero-mean and unit variance. Table 2 lists the parameters of our Att-CNN. In addition, we assume  $N_d$ , epoch, batch size and learning rate of 20, 100, 100 and 0.005, respectively. The following experiments are conducted in a PC which has a 3.8 GHz AMD-R5-2600 CPU, a GeForce RTX 2060 with a 6 GB frame buffer and 16 GB RAM.

### 4.1 Fair antenna power allocation

To validate our fair antennas power allocation method, we conducted experiments to compare the results of fair and unfair antennas power allocation. We assume that  $P_t = 1e - 2$  and  $\lambda_k = 1e - 4$ , and  $I_{th}$  varies from  $1e - 6$  to  $1e - 8$ . Figure 6 shows comparisons of fair and unfair power allocation over users and antenna elements. Three sub-boxes (sub-plots) are exhibited. Each sub-plot represents one experiment/comparison for a different configuration. The fair scheme (left) heat map is compared to the unfair scheme (right) in every sub-plot. Horizontal and vertical axes are the users and antennas, respectively. One pixel means allocated specific power on an antenna to a CR user. Color depth represents the allocated power value such that the closer to the dark blue is the bigger. It is obvious that the fair model assigns power to all users without focusing all power on particular indices. However, the unfair model focuses the power assignment on a single CU index.

### 4.2 Power allocation performance

As a typical benchmark, we implemented an equal power method (EPM), in which  $P_{k,i}$  can be calculated by

**Table 2** Att-CNN parameters

Name	Value
$N_t$	1
$L_t$	99
$N_c$	3
$N_{f1}$	2
$L_{f1}$	891
$N_{f2}$	2
$L_{f2}$	9
$N_{f3}$	2
$L_{f3}$	9

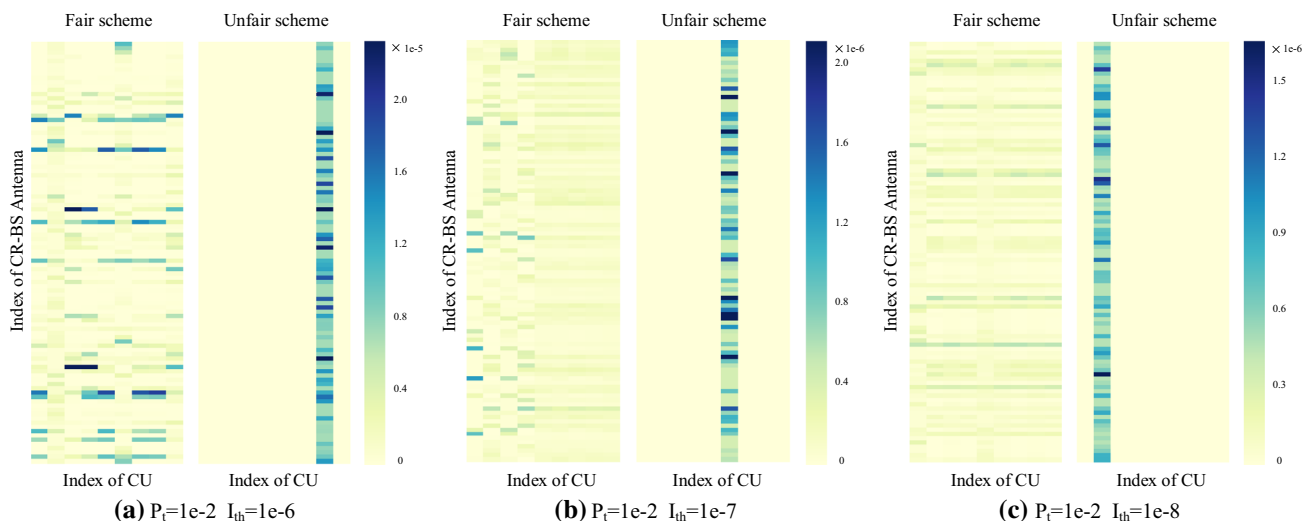


Fig. 6 Comparison of fair and unfair antennas power allocation with different parameters

$P_{k,i} = P_t/N_tK/I_d$ . Note that the division over  $I_d$  is meant for meeting the constraint  $C3$  in Eq. (3). Then, we implemented a FNN to validate the novelty of the proposed attention-based CNN, i.e., Att-CNN, from the perspective of neural network structures. Figure 7 shows the structure of the FNN. The data initialization network and results calculation network keep the same with the Att-CNN, and three ReLU layers with 891 neurons are used to replace the attention networks and convolutional layers.

To compare the performance of antennas power allocation methods, we defined two signal-to-noise ratios as follows.

$$\begin{cases} SNR_p \triangleq \frac{P_t}{\sigma^2}, \\ INR \triangleq \frac{I_{th}}{\sigma^2}. \end{cases} \quad (19)$$

Then we conducted two experiments using the EPM, Att-CNN and FNN, where we changed the  $SNR_p$  and  $INR$  from 0 to 50 dB to observe changes of SE. We assumed that  $\sigma^2 = 1e - 9$ ,  $P_{PR} = 1e - 4$  and  $\lambda_k = 0$  in these two experiments.

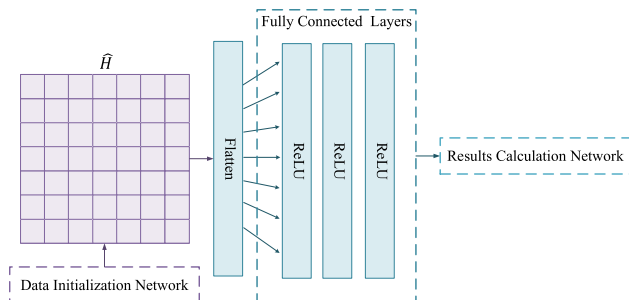


Fig. 7 A fully connected neural network which only replaces the attention and convolutional parts with three ReLU layers in Att-CNN

Figure 8 compares training effect between the Att-CNN and FNN. Around 10 epoch, both methods can get a good results. The FNN has a more serious overfitting than the Att-CNN. More parameters from FNN and limited training sets may cause this issue. After we added the attention mechanism and CNN into the network structure, the Att-CNN got better SE and reduced overfitting significantly.

Figure 9 shows the SE against  $SNR_p$ , when  $I_{th} = 1e - 6$ . The gain gap between Att-CNN and FNN is larger when the  $SNR_p$  increases, up to 0.588 b/s/Hz. The EPM is the worst which is even smaller than FNN by a gain gap up to 0.934 b/s/Hz.

Figure 10 shows the SE versus  $INR$ , when  $P_t = 1e - 4$ . It is clear that our proposed Att-CNN always outperforms than FNN and EPM when  $INR$  varies from 0 to 50 dB. The Att-CNN has a superiority over FNN and EPM, up to 0.596 and 1.586 b/s/Hz, respectively.

### 4.3 Computational performance

The proposed Att-CNN not only has better power allocation performance, but also has better computational

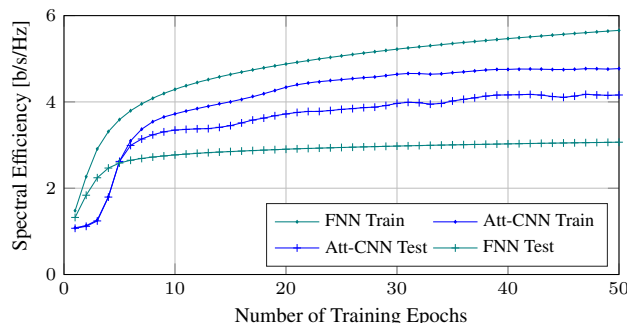


Fig. 8 Training and testing results of the Att-CNN and FNN



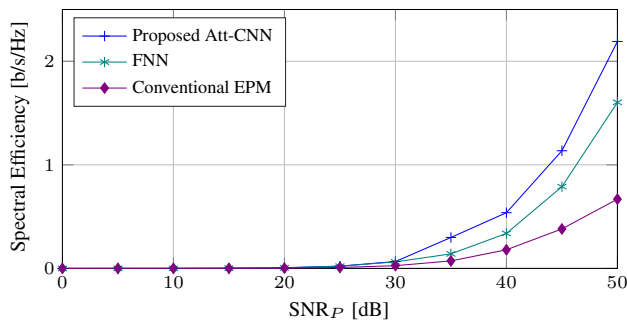


Fig. 9 SE versus  $SNR_P$  when  $I_{th} = 1e - 6$ ,  $P_{PR} = 1e - 4$  and  $\sigma = 1e - 9$

performance than the FNN. We use floating point operations, i.e., FLOPs, and number of parameters to evaluate computational performance.

Figure 11a shows FLOPs versus the number of users in the range between 2 and 9, and the number of antennas is fixed to 99. Only one case (i.e., the number of users is less than 3) is that the FNN has less FLOPs than Att-CNN. When there are more than three users, the FLOPs of the FNN increase sharply, while Att-CNN increases slowly. Figure 11b shows FLOPs against the number of antennas with 9 users. The Att-CNN has less neural network parameters than the FNN when the number of antennas is more than 49, which also increase slowly. Note that we assume our system as a massive MIMO system, hence cases with more antennas are more reasonable.

Figure 11c, d shows number of parameters against the number of users, and the number of antennas, respectively. They almost have the same trends with above FLOPs analysis.

## 5 Related work

### 5.1 Convolutional neural network

Deep learning has dramatically improved the novelties in many fields, such as speech recognition, visual object

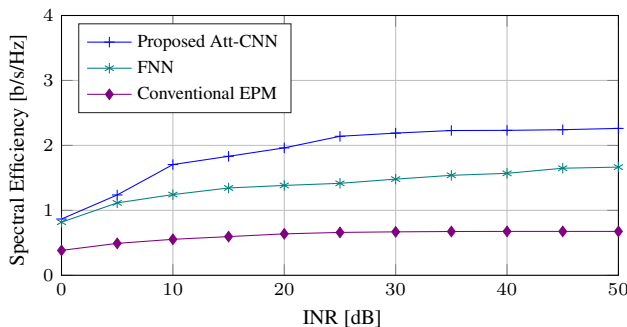


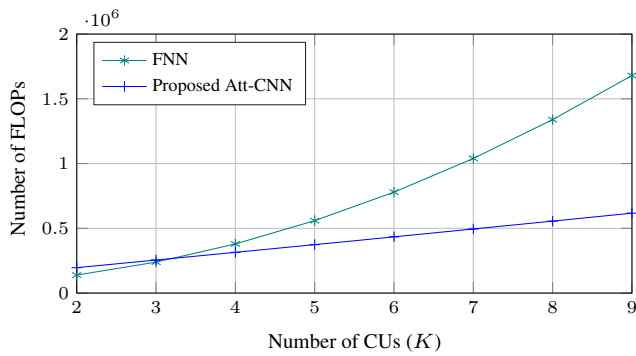
Fig. 10 SE versus  $INR$  when  $P_t = 1e - 4$ ,  $P_{PR} = 1e - 4$  and  $\sigma = 1e - 9$

recognition, object detection [25, 26]. Deep learning can be represented in different structures, e.g., fully connected neural network (FNN), recurrent neural network (RNN) and convolutional neural network (CNN). Among them, CNN have been widely studied in many fields.

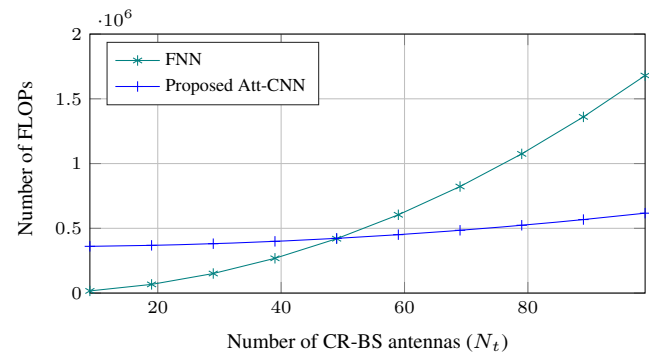
Fukushima [27] provided a pioneering research on CNN in 1980, wherein he proposed a neocognitron model concluded convolution and pooling layers. Lecun et al. [28] applied back propagation in their LeNet-5 which became the prototype of CNN. In 2012, Hinton et al. [29] improved the performance of CNN in image recognition with Alexnet which used deep structure and dropout method. Based on the previous researches, Lecun et al. [30] improved error rate to 11% by their Dropconnect. Later on, the error rate increased to 6.7% by Yan et al. [31] where they proposed a flexible CNN structure, called Network in Network. Besides image classification, examples on applications applied for CNN are: object detection [32], fault prediction [33], natural language processing [34] and so on. However, none of CNN researches about antennas power allocation are found.

### 5.2 Attention mechanism

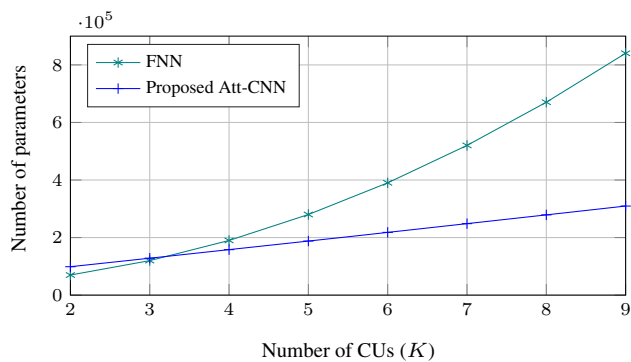
Sutskever et al. [35] proposed a sequence to sequence model in 2014, which has two problems: 1) When sentences are too long, the model performance will sharply decrease; 2) Different words in every sentence have the same priority. These problems are also mentioned in computer vision by Mnih et al. [36]. To address these issues, Bahdanau et al. [37] proposed attention mechanism, which implemented soft attention and provided some visual experimental figures. Later on, Xu et al. [38] applied the attention mechanism into computer vision, where they proposed two types of attention mechanisms: soft and hard attentions. In 2015, Luong et al. [39] improved these two attentions to the other two: local and global attentions. In 2017, Ahmed et al. [40] proposed a new network structure, named transformer which included a self-attention mechanism. The attention mechanisms have many applications. For instance, Li et al. [41] leveraged the attention mechanism to focus on the objects clicked by users in recommended systems based on session process; Liu et al. [42] proposed a gated multilingual attention framework to address the issue of data sparsity and monolingual ambiguity in event detection tasks; Du et al. [43] proposed a new self-attention mechanism which leveraged the relationships between local features and applied in some industrial video data. Min et al. [44] proposed bottom-up and top-down attention mechanism can assign higher weights to sensitive features, thereby limiting some redundant information. Up to authors' best knowledge, none of the state-of-the-art has introduced the attention



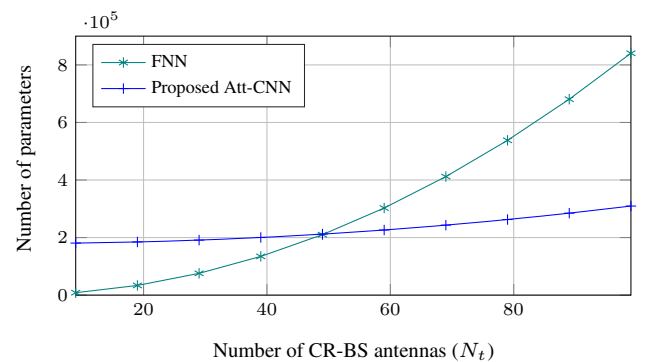
(a) FLOPs versus number of CR users.



(b) FLOPs against number of CR-BS antennas.



(c) Number of parameters versus number of CR users.



(d) Number of parameters against CR-BS antennas.

**Fig. 11** Complexity analysis for FNN and proposed Att-CNN

mechanism in the radio resource management context, particularly the fair power assignment task. The novel use of the attention mechanism in this work is motivated by its ability that allows for input features to dynamically come to the forefront as needed [38]. Such a characteristic is inspired from the related literature above.

### 5.3 Resource allocation

From the resource allocation perspective, machine learning methods have provided fast processing and can resolve optimization problems for time sensitive tasks. Those tasks are treated as a black box given that the relationship between the input and output is learnable via deep neural network [25]. Without loss of generality, deep neural network can approximate and solve non-convex optimizations known to be NP-hard [45]. In non-CR context, authors of [46] have developed an approximation for a weighted MMSE optimization by a deep neural network. The aim was to provide less computational complexity and hence real-time processing at almost similar performance. In the CR context, Zhou et al. [47] have implemented a deep neural network for resource allocation, namely spectral efficiency and energy efficiency maximizations. The training data have been obtained by specific conventional

strategies in the literature. Such dependency on conventional algorithms raises the computational complexity. Liu et al. [48] have employed weighted sum of CR interference power as objective for a minimization function with quality of service constraints for both CR and PR networks. Since the problem is non-convex, a message-passing algorithm based on deep learning has been utilized. A spectral efficiency maximization problem for a set of transceiver pairs has been solved by a FNN in [49]. Device-to-device communication links were considered and the FNN was trained by data set for which normalization has been conducted for timely efficiency. This model doesn't address infrastructure-based networks nor fairness constraints which are vital requirements in real networks.

## 6 Conclusion

The paper built a mathematical model for an unfair antennas power allocation issue in MIMO systems. Then, an attention-based convolutional neural network, i.e., Att-CNN, was proposed to address the issue. There, an  $h_0$  attention network was used to reevaluate the weights of channel gains in terms of  $h_0$ ; an  $h_k$  attention network was used to change the weights of the channel gains among

users; and a CNN was applied to decrease floating point operations (FLOPs) and number of network parameters. We used heap maps to compare fair and unfair allocation varying the interference thresholds, which verified the fair effect. To validate power allocation performance of our Att-CNN, we compared it with equal power allocation and a fully connected neural network (FNN) from the aspects of spectral efficiency against  $SNR_p$  and  $INR$ . At last, we analyzed the FLOPs and number of parameters of the Att-CNN and the FNN, which indicated the superiority of our Att-CNN.

In future research, we intend to further improve the computational performance of our Att-CNN and apply it in realistic industrial systems.

**Acknowledgements** This work was supported by a grant from the National Natural Science Foundation of China (No.U1609211).

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

- Wu J, Zhu X, Zhang C, Yu PS (2014) Bag constrained structure pattern mining for multi-graph classification. *IEEE Trans Knowl Data Eng* 26(10):2382–2396
- Wu J, Pan S, Zhu X, Cai Z (2015) Boosting for multi-graph classification. *IEEE Trans Cybern* 45(3):416–429
- Culbertson H, Kuchenbecker KJ (2017) Ungrounded haptic augmented reality system for displaying roughness and friction. *IEEE/ASME Trans Mechatronics* 22(4):1839–1849
- Mitola JI (2002) Cognitive radio: an integrated agent architecture for software defined radio. PhD Dissertation
- Hykin Simon (2005) Cognitive radio: brain-empowered wireless communications. *IEEE J Sel Areas Commun* 23(2):201–220
- Akyildiz IF, Lee WY, Vuran MC, Mohanty S (2008) A survey on spectrum management in cognitive radio networks. *IEEE Commun Mag* 46(4):40–8
- Bellalta Boris (2016) IEEE 802.11ax: high-efficiency WLANs. *IEEE Wirel Commun* 23(1):38–46
- 5G-Working-Groups. New radio: overall description. TS 38.300, 3GPP, 2019
- Lee K-J, Lee I (2011) Mmse based block diagonalization for cognitive radio mimo broadcast channels. *IEEE Trans Wireless Commun* 10(10):3139–3144
- Wang S, Ge M, Wang C (2013) Efficient resource allocation for cognitive radio networks with cooperative relays. *IEEE J Sel Areas Commun* 31(11):2432–2441
- Jindal N, Rhee W, Wishwanath S, Jafar SA, Goldsmith A (2005) Sum power iterative water-filling for multi-antenna gaussian broadcast channels. *IEEE Trans Inf Theory* 51(4):1570–1580
- Zhang R, Liang Y-C (2008) Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. *IEEE J Selected Topics Signal Process* 2(1):88–102
- Yaqot Abdullah, Hoehner Peter Adam (2017) Efficient resource allocation in cognitive networks. *IEEE Trans Veh Technol* 66(7):6349–6361
- Ban T-W, Lee W (2019) A deep learning based transmission algorithm for mobile device-to-device networks. *Electronics* 8(11):1361
- Lee M, Xiong Y, Yu G, Li G (2018) Deep neural networks for linear sum assignment problems. *IEEE Wireless Commun Lett* 7(6):962–965
- Wen C-K, Shih W-T, Jin S (2018) Deep learning for massive mimo CSI feedback. *IEEE Wireless Commun Lett* 7(5):748–751
- Ye H, Li G, Juang B-H (2018) Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Wireless Commun Lett* 7(1):114–117
- Joung J (2018) Machine learning-based antenna selection in wireless communications. *IEEE Commun Lett* 7(1):114–117
- Lin T, Zhu Y (2020) Beamforming design for large-scale antenna arrays using deep learning. *IEEE Wireless Commun Lett* 9(1):103–107
- Luo W, Li Y, Urtasun R, Zemel R (2016) Understanding the effective receptive field in deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 4898–4906
- Ngo H, Maryetta T, Larsson E (2011) Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models. In: *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP) 2011 May 22* (pp 3464–3467). IEEE.
- Wu J, Pan S, Zhu X, Zhang C, Wu X (2018) Multi-instance learning with discriminative bag mapping. *IEEE Trans Knowl Data Eng* 30(6):1065–1080
- Wu J, Hong Z, Pan S, Zhu X, Zhang C, Cai Z (2014) Multi-graph learning with positive and unlabeled bags. In: *Proceedings of the 2014 SIAM international conference on data mining*, pp 217–225. SIAM
- Wu J, Cai Z, Zeng S, Zhu X (2013) Artificial immune system for attribute weighted naive bayes classification. In: *The 2013 international joint conference on neural networks (IJCNN)*, pp 1–8
- LeCun Yann, Bengio Yoshua, Hinton Geoffrey (2015) Deep learning. *Nature* 521(7553):436
- Zhang Yongshan, Jia Wu, Zhou Chuan, Cai Zhihua (2017) Instance cloned extreme learning machine. *Pattern Recogn* 68:52–65
- Fukushima Kunihiko (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36(4):193–202
- LeCun Yann, Bottou Léon, Bengio Yoshua, Haffner Patrick et al (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems* (pp 1097–1105)
- Wan L, Zeiler M, Zhang S, Le Cun Y, Fergus R (2013) Regularization of neural networks using dropconnect. In: *International conference on machine learning* pp 1058–1066
- Lin M, Chen Q, Yan S (2013) Network in network. *arXiv preprint arXiv:1312.4400*
- Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, pp. 91–99
- Wen L, Li X, Gao L A transfer convolutional neural network for fault diagnosis based on resnet-50. *Neural Comput Appl*
- Li L, Goh TT, Jin D How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Comput Appl*
- Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In *Advances in NIPS*, 2014

36. Mnih V, Heess N, Graves A et al (2014) Recurrent models of visual attention. In: *Advances in neural information processing systems*, pp 2204–2212
37. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*
38. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: Neural image caption generation with visual attention. *Computer Science*, pp 2048–2057
39. Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*
40. Ahmed K, Keskar NS, Socher R (2017) Weighted transformer network for machine translation. *arXiv preprint arXiv:1711.02132*
41. Li J, Ren P, Chen Z, Ren Z, Lian T, Ma J (2017) Neural attentive session-based recommendation. In: *Proceedings of the 2017 ACM on conference on information and knowledge management*, pp 1419–1428. ACM
42. Liu Jian, Chen Yubo, Liu Kang, Zhao Jun (2018) Event detection via gated multilingual attention mechanism. In *Thirty-Second AAAI Conference on Artificial Intelligence*
43. Du Yang, Yuan Chunfeng, Li Bing, Zhao Lili, Li Yangxi, Hu Weiming (2018) Interaction-aware spatio-temporal pyramid attention networks for action classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 373–389
44. Xia Min, Liu Wan'an, Xu Yiqing, Wang Ke, Zhang Xu (2019) Dilated residual attention network for load disaggregation. *Neural Computing and Applications*
45. Luo Z-Q, Zhang S (2008) Dynamic spectrum management: Complexity and duality. *IEEE Journal of Selected Topics in Signal Processing* 2(1):57–73
46. Sun H, Chen X, Shi Q, Hong M, Fu ND, Sidirioulos X Learning to optimize: Training deep neural networks for interference management. *IEEE Transactions on Signal Processing*, 66(20):5438–5453
47. Zhou F, Zhang X, Hu RQ, Papathanassiou W, Meng A (2018) Resource allocation based on deep neural networks for cognitive radio networks. In *Proc. IEEE International Conference on Communications in China*, pages 40–45. IEEE, Aug
48. Liu Miao, Song Tiecheng, Jing Hu, Yang Jie, Gui Guan (2019) Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks. *IEEE Trans Veh Technol* 68(1):641–653
49. Lee Woongsup (2018) Resource allocation for multi-channel underlay cognitive radio network based on deep neural network. *IEEE Communication Letters* 22(9):1942–1945

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.